The SMART-I^2: A new approach for the design of immersive audio-visual environments

Marc Rébillat, Etienne Corteel, Brian F.G. Katz

Marc Rébillat[1, 3] – Étienne Corteel[2] – Brian F.G. Katz[1]

[1] Audio & Acoustic, LIMSI-CNRS, 91403 Orsay, France
[2] *sonic emotion*, CH-8154 Oberglatt, Switzerland
[3] LMS-CNRS, École Polytechnique, 91128 Palaiseau, France

The SMART-I²: A new approach for the design of immersive audio-visual environments.

Abstract

The SMART-I² aims at creating a precise and coherent virtual environment by providing users with both audio and visual accurate localization cues. Wave field synthesis, for audio rendering, and tracked passive stereoscopy, for visual rendering, individually permit high quality spatial immersion within an extended space. The proposed system combines these two rendering approaches through the use of large multi-actuator panels used as both loudspeaker arrays and as projection screens, providing a more perceptually consistent rendering.

Keywords: Wave field synthesis, Tracked passive stereoscopy, Multi-actuator panels

1. Introduction

In recent years, the advancement of immersive environments has separately produced systems with improved quality for 3D graphical rendering and also for 3D audio rendering. Despite these advances, only few audio-visual systems of high quality have been developed (Springer et al. 2006 or Kuhlen et al. 2007 for example). However, in any audio-visual (AV) application, the sensation of immersion and the intelligibility of the AV scenes depend highly on the quality of both the audio and the visual renderings (Blauert 2005). It is thus important to fulfill the requirements for both technologies in order to achieve a perceptually consistent rendering.

The SMART-I² (Spatial Multi-user Audio-visual Real-Time Interactive Interface), attempts to solve this problem by rendering spatially coherent audio and visual information. The sound rendering is realized using wave field synthesis (Berkhout et al. 1993) which relies on physical based reproduction of sound fields within an extended listening area. The graphical scene is rendered using tracked passive stereoscopy (Frölich et al. 2005), which presents users with the correct rendered visual image for each eye separately. These two technologies are combined together using two large multi-actuator panels (Boone 2004) which act both as projection screens and as loudspeaker arrays.

2. Wave Field Synthesis (WFS)

a. Physical basis

Wave Field Synthesis (WFS) is a spatialized sound rendering technology which was first developed at Delft University (Berkhout et al. 1993). It is an audio implementation of Huygens's principle, which states that *``Every sound field emerging from one primary sound source can be reproduced by summing* *contributions of an infinite and continuous distribution of secondary sound sources''.* At the theorical level, WFS allows one to synthesize a sound source at any given position. Implementations of WFS are simplified versions of this principle, typically using a linear array of equally spaced loudspeakers. Figure 1 illustrates the principle of WFS.

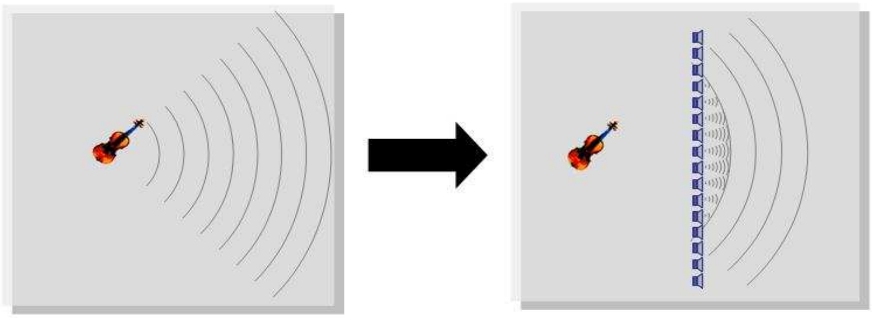

Figure 1: Illustration of sound rendering using WFS.

The violin on the left part is the primary source producing the target natural sound field. The linear array of secondary sound sources on the right produces, through summation of the contributions of each loudspeaker driven appropriately, a synthesized sound field equivalent to the original target field. The sound field of the virtual violin is then synthesized, and perceived by users in the reproduction area as emanating from the precise spatial location of the violin. Additional sound sources may be simultaneously synthesized through simple linear superposition.

b. Spatialized sound rendering

Using this physical basis, different types of fundamental sound sources, or sound fields, can be synthesized (see figure 2). Plane waves represent sound objects situated far away from the immersion area and are perceived as coming from a constant angle, independent of listener position. Point sources represent sound objects near the immersion area. Point sources can be synthesized at positions behind or in front of the loudspeaker array. Such focused sources (i.e. point sources in front of the array) are perceived as being physically present in the immersion area.

Sound rendering using WFS, due to its approach in physically recreating the entire field within a spatially large area, is not limited to a single user at a single location. The sound perspective, including parallax, is correct for every user in the immersion area, without the need of a tracking device.

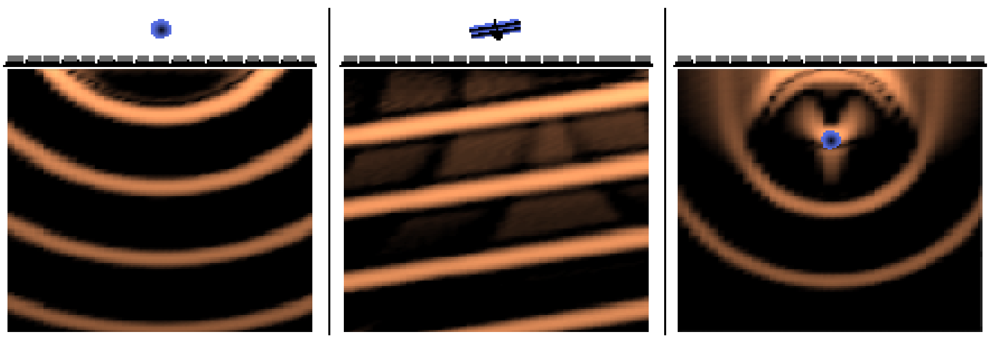

Figure 2: Synthesis of different wave types. Left: Point source behind the loudspeaker array. Center: Plane wave. Right: Point source in front of the loudspeaker array.

As previously stated, practical WFS implementations are limited to a linear array and hence reproduction is optimized for the horizontal plane. Auditory perception is the most precise and more stable in the horizontal plane and therefore a more pertinent choice for array orientation. With this restriction, which reduces the required calculation power, the digital audio processing can be done with a latency of less than 5 ms. this is more than sufficient for real time AV applications.

3. Design of the SMART-I²

This section provides a brief overview of the SMART-I2 system. A more complete technical description of the system can be found in Rébillat et al. 2008.

a. Tracked stereoscopy (TPS)

To produce a 3D visual rendering, each eye of the user must see the scene from a slightly different point of view. One means of realizing this is to use light polarization properties to independently address each eye of the user. The user wears special polarized glasses for visual cross-talk cancellation. The graphic rendering should also be adapted to the user's head position in order to maintain the correct point of view. Using this approach, the 3D visual rendering is coherent regardless of the user's position in the viewing area. This technique is referred as tracked passive stereoscopy (TPS).

b. Audio-visual integration with Multi-Actuator Panels (MAPs)

The integration of the two different technologies, TPS (see section 3.a) and WFS (see section 2), is achieved through an innovative use of multi-actuator panels (MAPs) (Boone 2004). MAPs are stiff lightweight panels with multiple electro-mechanical exciters attached to the backside. Typical MAPs multichannel loudspeakers are not larger than 1 m². For this project, a novel large dimension MAP has been designed, (i.e. 5 m² with a 4/3 ratio) in order to provide sufficient surface area and size to be used as a projection screen. To accommodate polarized light projection, the front face of the panel has been covered with metallic paint designed to preserve light polarization. Due to the nature of the MAPs design, screen displacements caused by acoustic vibrations are very small and do not disturb 3D video projection on the surface of the panel. Such a structure then allows one to efficiently integrate a 3D visual rendering technology and a spatialized sound rendering technology.

c. Architecture

The hardware architecture of the SMART-I² is schematically presented in figure 3. Two large MAPs of 2.6 m×2 m form a corner of stereoscopic screens and a 24 loudspeakers array. With this configuration, users can move within an immersion area of approximately 2.5 m×2.5 m. The rendering architecture of the SMART-I² is composed of three components. Virtual Choreographer is an open source real-time 3D graphics engine that relies on an XML-based definition of 3D scenes with graphic and sonic components. The WFS Engine is a real-time audio engine dedicated to low latency WFS rendering, capable of real-time filtering of up to 16 input channels (virtual sources) routed to 24 output channels (every exciter of the two MAPs). Max/MSP is a real-time audio analysis/synthesis engine using a graphical programming environment which, in addition to sound synthesis, provides peripheral interfacing, timing, and communications. An example of a simple AV scene rendered by the SMART-I² is given in figure 4.

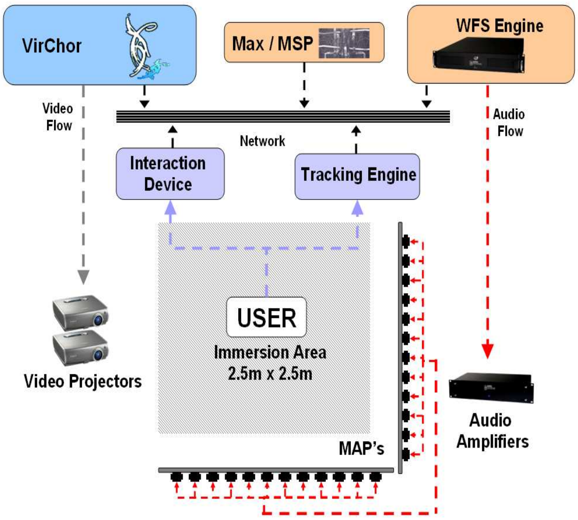

Figure 3: Global overview of the SMART-I²

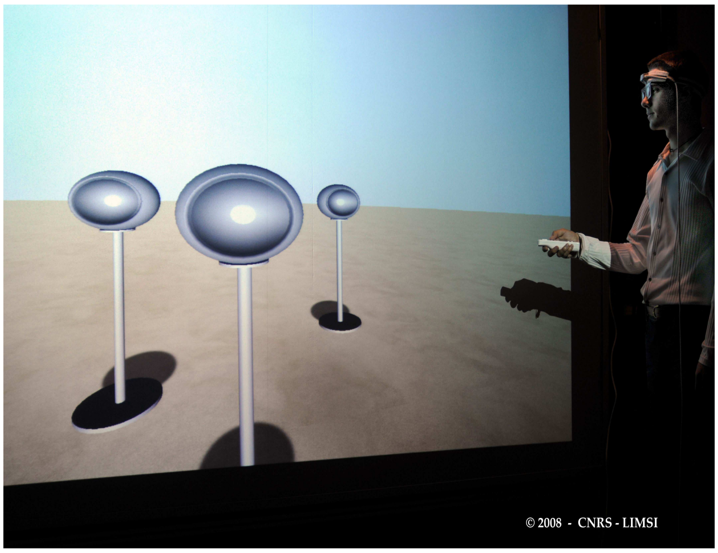

Figure 4: Example of AV scenes provided by the SMART-I²

4. Conclusion

The SMART-I² is a successful means of combining demands of high quality spatialized audio rendering and 3D video rendering through an innovative use of multi-actuators panels. The SMART-I² is thus a platform that can be used in a wide range of virtual reality applications.

Acknowledgements

The authors would like to thank Yves Maire, Daniel Lerin, Pascal Desroches, Rami Ajaj and Matthieu Courgeon for their precious help and advices.

References

Berkhout A.J., de Vries D. and Vogel P. (1993) Acoustic control by wave field synthesis. Journal of the Acoustical Society of America. Volume 93, no. 5, pp. 2764-2778.

Boone, M.M. (2004) Multi-Actuator panels (MAPs) as loudspeaker arrays for wave field synthesis. Journal of the Audio Engineering Society, Volume 52, no.7-8, pp. 712-723.

Blauert J. (2005) Communication Acoustics: Signal and communication technology. Springer-Verlag.

Frölich B., Blach R. and Stefani O. (2005) Implementing Multi-Viewer Stereo Displays. WSCG Conference Proceedings, Plzen, Czech Republic.

Kuhlen T., Assenmacher I., and Lentz T. (2007) A true spatial sound system for CAVE-like displays using four loudspeakers. 2nd International Conference on Virtual Reality, Beijing.

Rébillat M., Corteel E., Katz B.F.G. (2008) SMART-I²: Spatial Multi-user audio-visual real-time interactive interface. 125th Convention of the Audio Engineering Society, San Francisco.

Springer J.P., Sladeczek C., Scheffler M., Hochstrate J., Melchior F. and Frölich B. (2006) Combining Wave Field Synthesis and Multi-Viewer Stereo Displays. Proceedings of the IEEE conference on Virtual Reality, Washington, IEEE Computer Society, pp. 237-240.